SUPPLEMENTAL INFORMATION

Rapid and scalable characterization of CRISPR technologies using an *E. coli* cell-free transcription-translation system

Ryan Marshall[1]*, Colin S. Maxwell[2]*, Scott P. Collins[2], Thomas Jacobsen[2], Michelle L. Luo[2], Matthew B. Begemann[3], Benjamin N. Gray[3], Emma January[3], Anna Singer[3], Yonghua He[3], Chase L. Beisel[2]*, Vincent Noireaux[1]*

[1]School of Physics and Astronomy
University of Minnesota Minneapolis, MN, 55455

[2]Department of Chemical and Biomolecular Engineering
North Carolina State University, Raleigh, NC 27695

[3]Benson Hill Biosystems
St. Louis, MO 63132, USA

Table of Contents

Supplemental Figure Legends

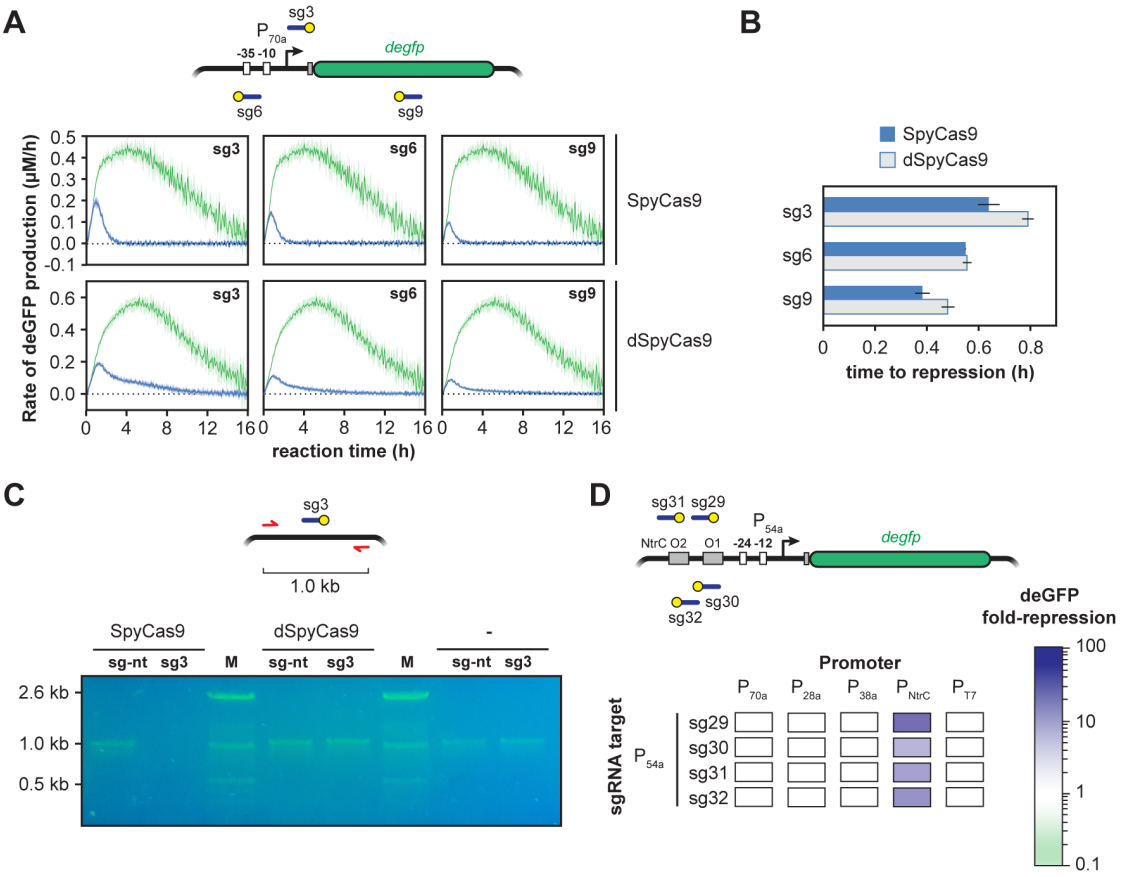

Figure S1. Related to Figure 1. The *S. pyogenes* Cas9 functions in TXTL. **A.** deGFP production rates from the time series curves from Figure 1B. Rates are shown for a targeting sgRNA (blue) and a non-targeting sgRNA (green). Rates are calculated by two-point numerical differentiation and smoothed with a five-point quadratic polynomial. The dark lines and light regions represent the average and S.E.M. of at least three runs. **B.** The number of hours before repression was observed in the time series curves from Figure 1B is shown. Values and error bars represent the average and S.E.M. of at least three replicates. **C.** 0.8% agarose gel showing cleavage of P70a-deGFP by SpyCas9, but not dCas9. A PCR of P70a-deGFP was performed after three hours of expression in the TXTL reaction, with primers flanking the g3 site in P70a-deGFP upstream by 197 bp and downstream by 877 bp. The full PCR product is 1074 bp. **D.** dSpyCas9-based repression by targeting NtrC binding sites. Top: Schematic of P54a-deGFP plasmid showing -24 and -12 consensus regions, NtrC binding sites and sgRNA target locations. Bottom: A matrix showing dSpyCas9-based repression using sgRNAs that target the NtrC binding sites. Values represent the mean of at least three repeated TXTL reactions.

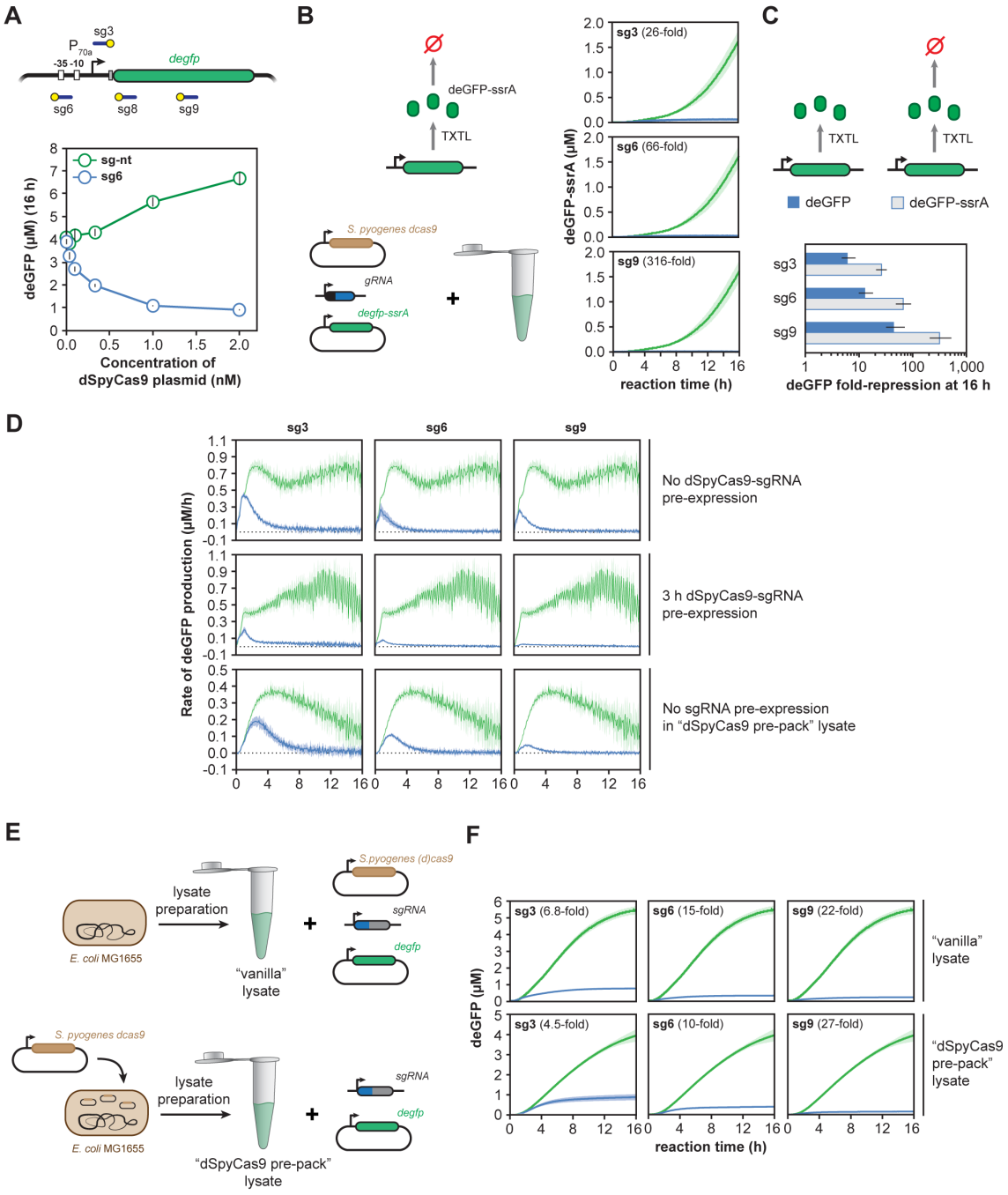

Figure S2. Related to Figure 2. Multiple factors affect dSpyCas9-based repression of reporter production in TXTL. **A.** Endpoint deGFP concentrations for TXTL reactions expressing the reporter plasmid, either a targeting (g6) sgRNA or a non-targeting (g-nt) sgRNA, and varying concentrations of the dCas9 plasmid. Values and error bars represent the average and S.E.M. of at least three replicates. **B.** Time course of deGFP-ssrA expression, where the reporter plasmid was targeted by a targeting sgRNA (blue) or a non-targeting sgRNA (green). The ssrA degron tag is recognized by the ClpXP protease that results in rapid turnover of the fusion protein. The dark lines and light

regions represent the average and S.E.M. of at least three runs. **C.** Fold-repression for reporter constructs encoding deGFP or deGFP-ssrA. Fold-repression is the ratio of deGFP concentrations after 16 hours of reaction for the non-targeting (green) over the targeting (blue) sgRNA. Error bars represent the S.E.M. from at least three repeated TXTL reactions. **D.** deGFP production rates from the time series curves from Figure 2B, as well as for TXTL reactions where dSpyCas9 was expressed in cells prior to generating the lysate. Rates are calculated by two-point numerical differentiation and smoothed with five-point quadratic polynomial. The dark lines and light regions represent the average and S.E.M. of at least five runs. **E.** Pre-expressing dSpyCas9 protein in *E. coli* prior to generating the TXTL lysate. A schematic for preparing either “vanilla” or “dSpyCas9 pre-pack” lysate is shown. **F.** Time series of deGFP concentration for vanilla or dSpyCas9 pre-pack cell-free reactions expressing either non-targeting sgRNA (green) or targeting sgRNAs (blue) is shown. The dark lines and light regions represent the average and S.E.M. of at least five runs.

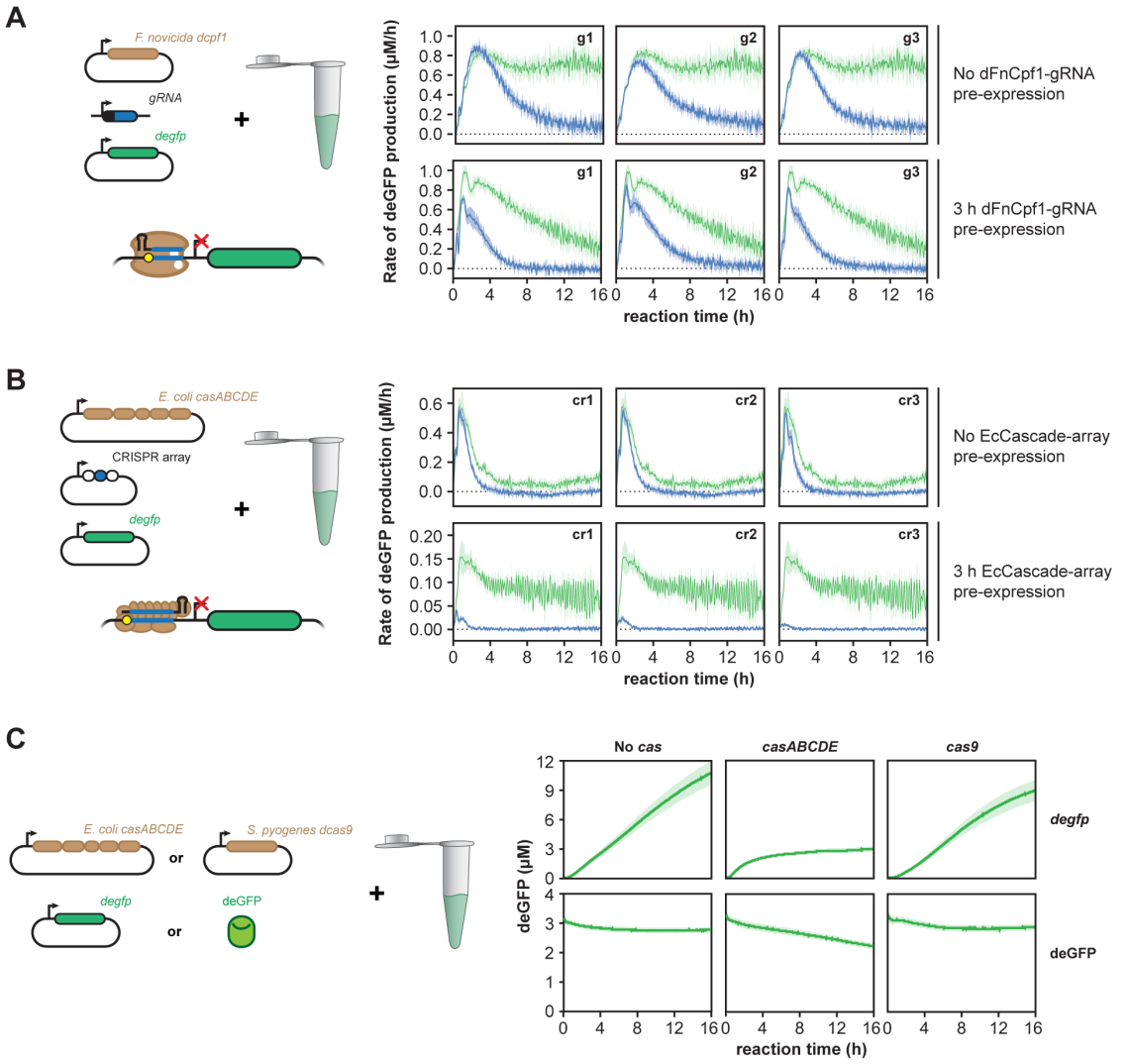

Figure S3. Related to Figure 4. Single effector and multi-protein effector Cas proteins function efficiently in TXTL. **A.** deGFP production rates from the time series curves from Figure 4A for a targeting guide RNA (blue) or a non-targeting guide RNA (green). Rates are calculated by two-point numerical differentiation and smoothed with five-point quadratic polynomial. The dark lines and light regions represent the average and S.E.M. of at least five runs. **B.** deGFP production rates from the time series curves from Figure 4B for a targeting guide RNA (blue) or a non-targeting guide RNA (green). Rates are calculated by two-point numerical differentiation and smoothed with five-point quadratic polynomial. The dark lines and light regions represent the average and S.E.M. of at least five runs. **C.** Time course of deGFP fluorescence in TXTL when deGFP is expressed from a plasmid (top) or added as purified recombinant protein (bottom). Reactions included the expression of EcCascade or the SpyCas9 or nothing (No *cas*). The dark lines and light regions represent the average and S.E.M. of at least six runs.

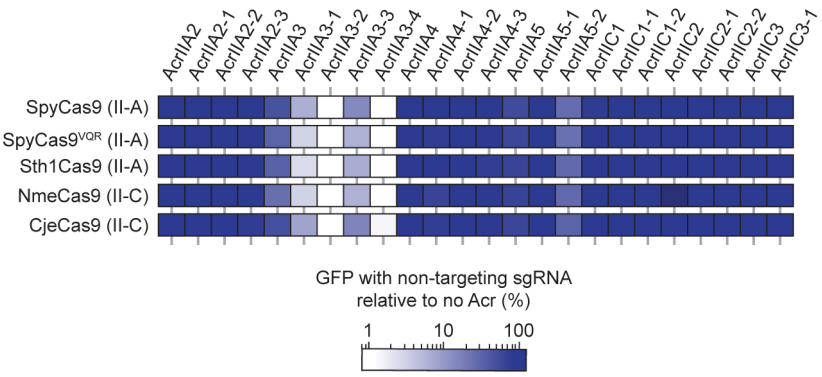

Figure S4. Related to Figure 5. Effect of anti-CRISPR proteins on expression of GFP in TXTL. The matrix shows the endpoint GFP expression for TXTL reactions with non-targeting sgRNA, the anti-CRISPR protein, and the Cas9. The values are reported in comparison to the same TXTL reaction without the anti-CRISPR protein. Endpoints were taken after 18 hours of incubation. Values represent the mean of at least three technical replicates.

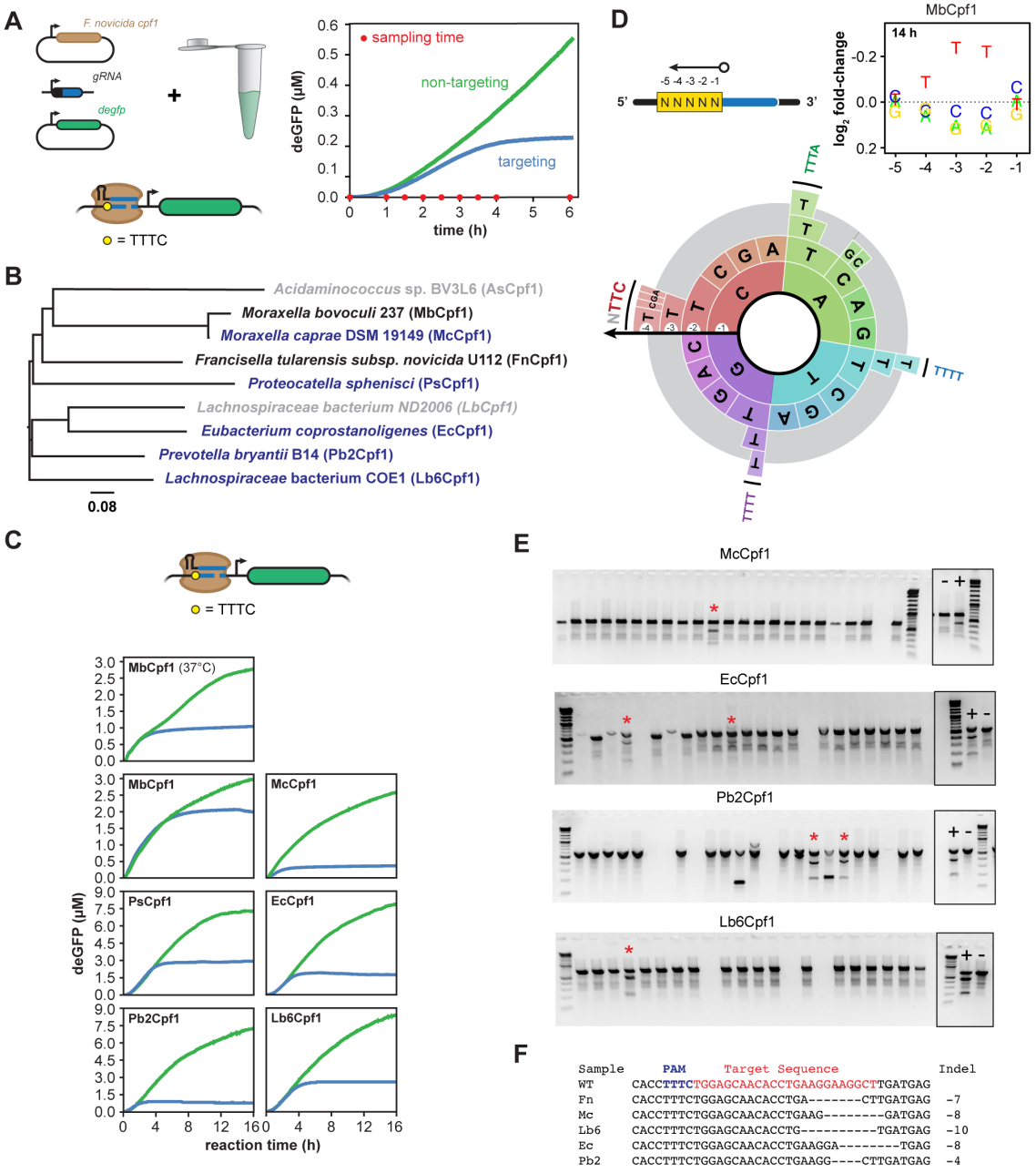

Figure S5. Related to Figure 6. Dynamics of DNA cleavage by FnCpf1. **A.** A time series is shown for deGFP concentration in TXTL reactions expressing FnCpf1, deGFP, and either a targeting guide RNA (g4) or non-targeting guide RNA (g-nt). Red dots show times where samples were collected in a parallel reaction expressing FnCpf1, a targeting guide RNA, and containing a 5N PAM library as part of the PAM determination assay (Figure 6A). **B.** Phylogenetic tree of Cpf1 nucleases that were previously uncharacterized (blue), previously characterized but also tested in this work (black), or previously characterized and not explicitly tested in this work (gray). Note that the PAM for MbCpf1 was previously reported in Zetsche et. al. *Cell* 2015, although the nuclease was not subsequently characterized. The tree was constructed using MUSCLE based on

the amino acid sequence of each nuclease. **C.** Time courses for TXTL reactions expressing a Cpf1 nuclease from a plasmid, a targeting or non-targeting sgRNA from linear DNA, and the deGFP reporter from a plasmid. In all cases, the target site was flanked by a 5' TTTC PAM. **D.** Results from the TXTL-based PAM determination conducted on MbCpf1. The assay was conducted at 37ºC with an incubation time of 14 hours. Note that depletion was limited compared to the other tested Cpf1 nucleases, resulting in a different scale for the vertical axis as compared to those in Figures 6B,E. Results of the assay are displayed as the fold-change of each nt at each position of the PAM library in comparison to the original PAM library (top) and the PAM wheel showing the determined PAM sequences (bottom). **E.** Representative T7EI assays screening rice callus for indels. Each lane corresponds to a separate callus piece, while + and – lanes correspond to positive controls (WT DNA mixed with a gBlock containing a defined mutation) and negative controls (WT DNA), respectively. Asterisks indicate samples that showed a T7EI signal indicative of an indel; these samples were further analyzed by Sanger sequencing to validate genome editing events. **F.** Genome editing events mediated by Cpf1 nucleases in rice. Representative deletions are shown that were produced by Lb, Mc, Lb6, and Ec Cpf1 nucleases and determined by Sanger sequencing. Dashes indicate deletions relative to the wild-type sequence.

Supplemental Table Legends

Supplemental Table S1. Related to Figures 1 - 6 and S1 - S5. Plasmid, gBlocks, bacterial cell stocks, primers, and Cpf1 sequences used in this study.

Supplemental Data

Supplemental Data S1. Related to Figure 6. Krona plots used to generate the PAM wheels. The plots capture the output of the TXTL-based PAM determination assays conducted on FnCpf1, Lb6Cpf1, Pb2Cpf1, EcCpf1, PsCpf1, McCpf1, and MbCpf1.

Supplemental Protocols

Supplemental Protocol S1. Related to Figure 1. Protocol for measuring the activity of CRISPR nucleases using TXTL.

Supplemental Protocol S2. Related to Figure 6. Protocol for performing the TXTL-based PAM determination assay.